REMARKS

The Office Action of May 14, 2004 presents the examination of claims 1, 4, 6, 7, 9, 11-13, 15-18, 30-36, 40, 41 and 43-45. Claims 6 and 43 are deemed allowable.

By the present amendment, claims 1-5, 7-42 and 44-45 are now canceled. New claims 46-77 are presented for examination.

Support for the new claims

New claims 46-77 find support in the specification in the Sequence Listing (sequence recitations), the paragraph bridging pages 14-15 (activity of the enzyme), page 16, lines 6-9 (hybridization conditions) and pp. 50-57 (product by process recitations).

Rejection for alleged lack of utility

Claims 9, 12, 13, 17, 18 and 44 stand rejected under 35 U.S.C. §§ 101 and 112, first paragraph, for alleged lack of proof of an asserted utility. These claims have been canceled, rendering this rejection moot. Applicants submit that this rejection should not be applied to the now pending claims.

The Examiner takes a position that there is no evidence of record that the claimed nucleic acids encode a protein having raffinose synthase activity. The Examiner asserts that mere homology to a known raffinose synthase gene is not sufficient,

as there are other genes known that have higher degrees of homology to some raffinose synthase gene, yet encode proteins having other activities. In particular, the Examiner has compared the *S. affinis* stachyose synthase to the *S. cuminis* raffinose synthase and found a degree of homology of 50% and with the *P. sativum* raffinose synthase and found a degree of homology of 51%. Applicants note that in their prior response, a sequence comparison was provided as Table 2; these same comparisons are in Table 2 and a degree of homology of only 43% is found for both.

The discrepancy in the degree of "homology" in the data set provided by the Applicant and by the Examiner is due to differences in the computer program used to analyze the data. Table 2 of Applicants' previous response shows overall sequence homologies(%) between raffinose synthases (RFSs), imbibition protein (SIP) and stachyose synthases (STSs). Applicants' homology data were calculated using Global Alignment (the alignment of sequences over their entire length) produced by the CLUSTAL sequence analysis program. The CLUSTAL program uses the algorithm of Wilbur and Lipman (see the attached Exhibit 1, a description of sequence analysis programs found at http://www.rfcgr.mrc.ac.uk/embnet.news/vol2_1/align.html).

On the other hand, the Examiner's analysis was conducted using the BLAST program, which is a Basic Local Alignment Search Tool. The BLAST program uses BLAST algorithm and makes local alignments (the alignment of some portion of two sequences) in order to search similarities of sequences. BLAST is a local alignment program, and does not make global alignments between sequences to calculate total percent homologies. (See, page 3, the 9th paragraph of attached Exhibit 2, http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.shtml). Moreover, "identities" values output in a BLAST search report are different from "homology" values. That is, "homology" means "similarity attributed to descent from a common ancestor", while "identity" means "the extent to which two sequences are invariant" (See, pages 2-3 of attached Exhibit 3, http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/glossary2.html). Thus, different values and scores calculated using different algorithms are based on different standards or criteria, and therefore it is not reasonable to discuss homology and similarity of sequences only by directly comparing such different values and scores.

The attached Table 3 shows the identities obtained using a similarity search using the BLAST program for the amino acid sequences of RFSs, SIP and STSs shown in Table 1 (the same as attached to Applicants' previous response). Among Sc-02, Sc-03,

Sc-04 and Sc-05, the identities were obtained by searching a patent database with default parameters, using the amino acid sequence of each protein as the "query", and using "Protein query vs. translated database (tblastn)" of the NCBI BLAST program. Also, other identities were obtained by searching the non-redundant database with default parameters, using the amino acid sequence of each protein as the "query", and using "Protein-protein BLAST (blastp)" of the NCBI BLAST program.

The identities between RFSs and SIPs are about 40%. The identities between RFSs and STSs range from about 40% to about 50%. On the other hand, the identities among RFSs are 60% or more. The identities among STSs are also 60% or more. That is, the identities among RFSs or the identities among STSs are higher than the identities between RFSs and SIPs or the identities between RFSs and STSs. Thus, based on the results of analyses by BLAST program, RFSs, SIPs or STSs can be distinguished. Applicants note that the conclusion reached from this analysis is consistent with the conclusion reached using the CLUSTAL analysis provided in their previous paper.

Applicants submit that the preponderance of the evidence of record establishes that Raffinose Synthases, Stachyose Synthases and Imbibation Proteins are, as groups, more homologous to each other they are among different synthase types. That is, any given Raffinose Synthase will have a greater degree of overall

sequence identity to another Raffinose Synthase than to a Stachyose Synthase or an Imbibation Protein. Accordingly, identification of a protein by homology analysis as having higher similarity to a Raffinose Synthase than to a Stachyose Synthase or Imbibation Protein is sufficient to establish that protein may be used in the manner similar to that in which know Raffinose Synthase proteins may be used.

Accordingly, the utility of the instantly claimed invention is established and the rejection of claims 9, 12, 13, 17, 18 and 44 stand rejected under 35 U.S.C. §§ 101 and 112, first paragraph, for alleged lack of proof of an asserted utility should not be applied to the present claims.

Rejection for lack of written description

Claims 1, 4, 7, 9, 11-13, 15-18, 30-36, 40 and 41 stand rejected under 35 U.S.C. § 112, first paragraph, for alleged lack of adequate written description of the invention. These claims have been canceled, rendering this rejection moot. Applicants submit that this rejection should not be applied to the present claims.

The Examiner takes a position that the specification provides no description of any generic structural feature that confers raffinose synthase activity upon a protein. The Examiner continues to rely upon the *University of California v.*

Eli Lilly case. Applicants have already made arguments distinguishing the facts of the present case from those of *Eli Lilly*. The Examiner asserts that the specification fails to describe any complete coding sequence of a raffinose synthase protein other than SEQ ID NO: 1 (or the complete amino acid sequence of SEQ ID NO: 2). This assertion by the Examiner is simply not correct. SEQ ID NO: 3 present a sequence of a complete raffinose synthase protein. The Examiner seems to know this, as he points only to SEQ ID Nos: 5 and 7 as not being complete sequences. Thus, the primary premise underlying the Examiner's position is not consistent with the actual facts.

Second, the Examiner asserts that the specification is not sufficient to meet Applicants' burden of establishing that the other nucleic acid and amino acid sequences described are actually raffinose synthase genes and proteins, respectively. This argument has been rebutted by the data and explanation thereof provided above.

The rejection has insufficient legal basis for the above reasons and this alone urges that it should not be applied to the present claims.

Moreover, many of the present claims in fact recite structural features that are plainly set forth in the Sequence Listing and distinguish the generic invention as claimed from other nucleic acids, and also describe functional outcomes (a

biological activity of the enzyme) that are associated with those structural features. The remaining claims describe the invention in product-by-process terms. Such a manner of claiming a generic invention is entirely proper. See, *Enzo Biochem, Inc. v. Gen-Probe, Inc.*, 63 USPQ2d 1609 (Fed. Cir. 2002). Accordingly, it is inappropriate to apply the instant rejection to the presently-pending claims.

Rejection for alleged lack of enablement

Claims 1, 4, 7, 9, 11-13, 15-18, 30-36, 40, 41 and 44 stand rejected under 35 U.S.C. § 112, first paragraph, for alleged lack of enablement by the specification. These claims have been canceled, rendering this rejection moot. Applicants submit that this rejection should not be applied to the present claims.

Applicants note first that analysis of enablement is a question of whether "undue experimentation" is required to practice the invention throughout its claim scope. Consideration of the question of undue experimentation is by weighing of each of several factors enumerated in *In re Wands*, 8 USPQ2d 1400 (Fed. Cir. 1988).

The Examiner fails to meet his burden of establishing a *prima facie* lack of enablement. The Examiner's analysis of the question of undue experimentation looks only at the factor of whether

working examples of the claimed invention are described in the specification and an assertion that it is unpredictable whether any particular nucleic acid produced according to the teachings of the invention would in fact exhibit raffinose synthase activity. This analysis is legally insufficient to establish *prima facie* lack of enablement, as the Examiner fails to consider the breadth of the claims, the nature of the invention, the level of ordinary skill in the art, the quantity of the experimentation needed, the guidance provided by the specification (other than the presence or absence of working examples) and the state of the art at the time the invention was made. Furthermore, the kind of predictability, *a priori* knowledge of functionality of the enzyme obtained using the methods of the invention, is not the kind of predictability envisioned by the Court in *Wands*. The instant rejection cannot properly be sustained against any claims.

The nature of the invention and the breadth of the claims

The claimed invention relates to isolated nucleic acids that encode an enzyme having a defined biological activity. The claims recite that the invention lies in a nucleic acid that is defined by inclusion of at least certain sequence features, hybridizes to a certain reference sequence and encodes a protein having a defined enzymatic activity.

The art of molecular biology, in particular the art of expression of recombinant proteins, is one in which the artisan of ordinary skill expects to perform a few weeks or months of experimentation in generating variants of a protein, then isolating clones encoding those variants and then (perhaps) re-cloning the isolated variants into vectors for expressing a protein, and then screening expressed proteins for activity.

The level of ordinary skill in the art

The artisan of ordinary skill in the art of cloning and expressing recombinant proteins is generally accepted as one having a Ph.D. degree and perhaps higher. Such a person is skilled in the design and performing of experiments for isolating DNA clones and for screening them for a desired property, for example encoding a protein having a particular activity.

The amount of experimentation needed

The amount of experimentation needed to practice the present invention is not unduly large or burdensome. The practitioner must isolate a template genomic DNA from an organism, perform a polymerase chain reaction using primers described in the specification to generate an amplified fragment, clone that fragment into an expression vector, express the encoded protein and then screen the protein for activity as a raffinose synthase.

All of these steps are either well-known in the art or described in detail in the specification and furthermore are expected to be performed by the artisan of ordinary skill.

The state of the art at the time the invention was made

At the time the invention was made, the state of the art of molecular biology was such that the various laboratory operations that must be performed to carry out the experimentation required to practice the instant invention, i.e. cloning of DNA molecules and expressing them in a host cell, were routine. Also, polymerase chain reaction amplification of nucleic acids was routine.

The raffinose content of a number of organisms, especially including plants and some algae, was known. The biochemistry of raffinose synthesis in plants had been established, and the role of raffinose synthases as rate-limiting of raffinose production was known.

A biochemical assay for raffinose synthase activity was described. See Exhibit 4 attached, Lehle et al., *Eur. J. Biochem.* 38:103 (1973).

The guidance provided by the specification including the presence or absence of working examples

The specification provides ample guidance to the skilled artisan for practicing the invention broadly. In particular, the

specification discloses in detail how to clone DNAs encoding putative raffinose synthase enzymes. The specification provides details such as organisms likely to be useful for isolating template genomic DNA or cDNA (see, e.g. page 1, lines 9-14) and methods for cloning DNA encoding a putative raffinose synthase enzyme from an RNA fraction, including an extensive list of primers that can be utilized for PCR amplification from templates obtained from different organisms (see, e.g. page 10, line 11 to page 18, line 14). The specification describes methods for expressing the cloned DNA in plant cells and in bacteria (see, e.g. page 24, line 3 to page 27, line 23) and an example of expression in bacteria (Example 8 beginning at page 39). The specification describes how to purify raffinose synthase from plant cells (see, e.g. Example 3 beginning on page 32). The specification describes a biochemical assay for raffinose synthase, referring to the Lehle article noted above and summarizing the procedure in Example 2 beginning at page 31.

The specification also provides a number of working examples of isolation of partial or complete raffinose synthase genes from a number of different plants. See, Examples 7 and 9 to 11) and of transformation of a plant (soybean) with a cloned DNA encoding a raffinose synthase (Example 13).

The predictability in the art

The Examiner asserts that the art of recombinant DNA cloning and recombinant protein expression is unpredictable. The Examiner argues that a practitioner of the invention must engage in trial and error experimentation to identify cloned DNAs that encode functional raffinose synthase genes.

The Examiner's argument is simply incorrect. First, the skilled artisan can follow detailed teachings in the specification of how to clone, express and evaluate DNAs that are likely to encode functional raffinose synthase enzymes. It is true that it is unpredictable whether any individual clone made in an experiment will include a DNA encoding a functional enzyme, but it is not unpredictable whether the skilled artisan would succeed in identifying at least one functional DNA in an experiment as a whole. To the contrary, it is very likely that the skilled artisan would find a cloned DNA encoding a functional enzyme by following the teachings of the specification.

The Examiner is urged to read the *Wands* case in detail. In that case, an invention related to isolation of hybridomas that secreted a particular antibody was deemed broadly enabled despite that extensive screening of many cloned cell lines was necessary AND that the success rate of the screening was only 2.8%, including experiments that failed to generate any operable clones at all. The *Wands* panel expressly stated that experimentation,

such as the cloning and screening experiments described in the present application, that is expected to be performed by the artisan of ordinary skill, is not undue experimentation.

Applicants submit that a proper weighing of the *Wands* factors will lead the Examiner to a proper conclusion that no undue experimentation is required to practice the present invention broadly. Accordingly, the instant rejection should not be applied against the present claims.

Rejection for obviousness-type double patenting

The Examiner is maintaining the provisional obviousness-type double patenting rejection of record. Applicants again request that the Examiner hold this rejection in abeyance until either this application or the '766 application is allowed, at which time an appropriate response in the form of either arguments distinguishing the invention or a terminal disclaimer will be filed in the application that remains under examination.

Applicants respectfully submit that the above remarks and/or amendments fully address and overcome the rejections of record. The present application is in condition for allowance. The Examiner is respectfully requested to issue a Notice of Allowance indicating that claims 6, 43 and 46-77 are allowed.

Should there be any outstanding matters that need to be

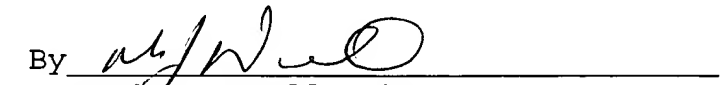
resolved in the present application, the Examiner is respectfully requested to contact Mark J. Nuell (Reg. 36,623) at the telephone number of the undersigned below.

Pursuant to the provisions of 37 C.F.R. §§ 1.17 and 1.136(a), the Applicants have petitioned for an extension of three (3) months to November 14, 2004, in which to file a reply to the Office Action, in the accompanying Request for Continued Examination. The required fee of $980.00 is enclosed therewith.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. § 1.16 or under 37 C.F.R. § 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By 
Mark J. Nuell, Ph.D.

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

DRN/jao
0020-4348P

Attachments: Table 1 (as per previous Amendment)
Table 3
Exhibits 1-4

Table 1

Code	Protein*	Organism	Accession**	Reference	Author/Assignee
Sc-03	RFS	***Beta vulgaris***	E37133	09/301,766	Sumitomo Chemical
Sc-05	RFS	***Brassica juncea***	E36417	09/301,766	Sumitomo Chemical
Sc-02	RFS	***Vicia faba***	E24423	08/992,914	Sumitomo Chemical
Sc-04	RFS	***Glycine max***		08/992,914	Sumitomo Chemical
Aj-05	RFS	***Cucumis sativus***	AF073744	Family GH36***	Ohsumi et al.
PsRFS	RFS	***Pisum sativum***	AJ426475	Family GH36	Peterbauer et al.
HvSIP	SIP	***Hordeum vulgare***	M77475	Family GH36	Heck et al.
PsSTS-1	STS	***Pisum sativum***	AJ311087	Family GH36	Peterbauer et al.
PsSTS-2	STS	***Pisum sativum***	AJ512932	Family GH36	Peterbauer et al.
VaSTS	STS	***Vigna angularis***	Y19024	Family GH36	Peterbauer et al.
AmSTS	STS	***Alonsoa meridionalis***	AJ487030	Family GH36	Voitsekhovskaja
SaSTS	STS	***Stachys affinis***	AJ344091	Family GH36	Pesch and Schmitz

*Protein: RFS, Raffinose synthase; SIP, Seed Imbibition Protein; STS, Stachyose synthase.

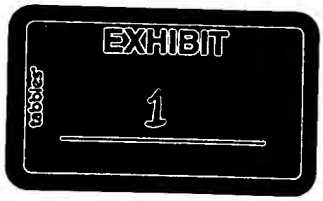
**Accession: GenBank Accession Number.

***Family GH36: glycoside hydrolase family 36 (see Carbohydrate-Active Enzymes (CAZy) database: http://afmb.cnrs-mrs.fr/CAZY/GH_36.html)

Table 3

Code	Sc-02	Sc-03	Sc-04	Sc-05	PsRFS	Aj-05	HvSIP	AmSTS	PsSTS-1	PsSTS-2	SaSTS	VaSTS
amino acids	800	783	781	777	798	784	757	868	853	847	863	857
Sc-02		61	75	61	88	63	41	52	50	50	49	50
Sc-03			65	63	62	66	40	49	39	39	49	49
Sc-04				67	75	65	38	39	50	43	51	52
Sc-05					60	70	40	51	51	51	52	50
PsRFS						65	40	51	49	49	51	50
Aj-05							39	49	50	50	50	50
HvSIP								38	38	38	38	38
AmSTS									63	64	81	62
PsSTS-1										96	63	72
PsSTS-2											64	72
SaSTS												64
VaSTS												

ALGORITHMS FOR MULTIPLE SEQUENCE ALIGNMENTS



Guy Bottu,
BEN, The Belgian EMBnet node.

Introduction.

In a previous issue of embnet.news, we considered the alignment of pairs of sequences and the search for similar sequences in databanks. We now turn our attention to multiple sequence alignments.

If you have several similar nucleic acid or protein sequences it is often useful to align corresponding bases or amino acids in columns. For instance, you might wish to group bases or amino acids that occupy similar positions in the three-dimensional structure which exercise similar functions or that have evolved by substitution from the same base or amino acid in an ancestral sequence. In the latter case you might also like to construct a phylogenetic tree.

1. Global alignments.

The Needleman and Wunsch algorithm for finding the best global alignment of two sequences can readily be extended to multiple sequences. The problem is that the time the computer needs for such a job is roughly proportional to the product of the sequence lengths. So, if aligning two sequences of 300 positions takes 1 second, aligning 3 sequences takes 300 seconds and aligning 10 sequences would take 300^8 seconds, which is longer than the lifetime of the universe!

Since searching for a best global alignment using a rigorous algorithm is not realistic for more than three sequences, a number of strategies have been developed to carry out a multiple global alignment in a reasonable amount of time with a reasonable chance of finding the best alignment. The GCG program pileup first aligns all possible pairs of sequences according to Needleman and Wunsch (for n sequences, this makes n*(n-1)/2 alignments). Then it uses the pairwise similarity scores to construct a tree using the UPGMA method (see below). Finally, this tree serves as a guide for a progressive multiple alignment starting from the tips. Once two sequences have been aligned, their relative alignment is no longer changed. Clusters of previously aligned sequences are treated as a linearly weighted profile when they are subsequently aligned with another sequence or another cluster.

Other approaches include:

- The very popular CLUSTAL program differs only from pileup in that it performs the initial pairwise alignments using the fast algorithm of Wilbur and Lipman. CABIOS 8:189 (1992). References
  you can obtain versions of CLUSTAL for UNIX and for VAX
- Starting with a search for words of n bases or amino acids that are common between the sequences. An example is Martin Vingron's program MALI. CABIOS 5:115 (1989). References.
  MALI is not distributed freely but may be obtained from its author Martin Vingron (vingron@embl-heidelberg.de)
- PIMA uses pattern-matching, rather then profile matching, while making the progressive alignment. PNAS 87:118 (1990)
  References
  PIMA can be obtained for UNIX and for VAX
- Building a phylogenetic tree, using a more elaborate algorithm, as the sequences are progressively aligned. An example is Jotun Hein's program TreeAlign. Meth.Enzymol. 18:626(1990)
  TreeAlign can be obtained for UNIX and VMS from the same address as given for Clustal (see above)
- Making the best multiple alignment in a limited area of alignment space. This can only realistically be performed with eight to ten sequences.

2. Local alignments.

There are cases where sequences share a similar region but are otherwise completely different. Take, for example, the amino acids in the active site of an enzyme or transcription factor binding sites in a DNA sequence. To handle these cases local multiple alignment algorithms have been developed. Usually they only look for ungapped alignments thereby avoiding the problem of choosing the optimal gap penalty. Two such programs have been developed at the NCBI :
MACAW by Schuler, Altschul and Lipman first tries to find high scoring segment pairs (HSPs) for each possible pair of sequences using the BLAST algorithm (with the sensitivity set high). It then assembles overlapping HSPs into blocks. An interesting feature of MACAW is that it does not try to align all sequences, but can pick out only those that share similar regions. Proteins 9:180 (1991). References
There are versions of MACAW obtainable for the PC under Windows and for the Mac.
The MACAW distribution also contains Gibbs (see below) and a pattern searcher.

The Gibbs sampler algorithm involves iteratively making a profile with stretches of n bases or amino acids, selected from the sequences, and then searches this profile against one of the sequences. The result of the search is used to weight the selection of the stretches at the next run. A drawback is that the user must choose the width n and the number of elements in each sequence and thus must have a certain idea of the outcome, or run the program several times. An interesting feature is that the Gibbs sampler algorithm avoids the choice of an externally added scoring scheme since it derives the highest scoring profile, in a

self-consistent manner, from the data. Science 262:208 (1992). References .
Gibbs for is available for UNIX.

3. blast3.

It is also worth mentioning the program blast3. This searches a protein against a protein databank using the BLAST algorithm (with the sensitivity set high) and then makes threefold alignments between the query sequence and each possible pair of databank sequences that have been found. Only the statistically significant threefold alignments which are made from three nonsignificant pairwise alignments are retained. blast3 is useful in finding proteins that share a region of only weak similarity. Occasionally it can show that a query sequence makes the bridge between two databank sequences whose relationship had not yet been suspected.
You can look at the Manual.
It is possible to access a BLAST (including blast3) server at the NCBI, either through WWW or with a specific blast Internet client that you can install on your computer. More INFO is available.

4. phylogenetic trees.

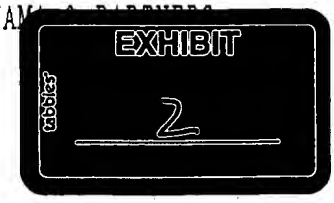
Ideally a researcher would like to have a black box in which to throw sequences and get out a fully annotated phylogenetic tree. This is, however, not possible for two reasons. First, an algorithm that considers all possible multiple sequence alignments and then, for each alignment, all possible phylogenetic trees and picks out the best one, would take too much time. That is why most phylogenetic programs work on previously aligned sequences. Second, the result is always strongly influenced by the criteria that are used to define the best tree. Phylogenetic analysis will be the subject of a separate column in a later issue of embnet.news. However, a few remarks seem appropriate here. There are three main kinds of tree building methods: distance matrix, maximum likelihood and parsimony.

Distance matrix methods first estimate the pairwise distances between the sequences (which means that the information in the alignment of two sequences is reduced to one number) while the other methods construct many trees from all the information in the multiple alignment and decide which is best.

The simplest distance based method is UPGMA (unweighted pair-group method using arithmetic averages) which involves iteratively taking together the two sequences that have the shortest distance from each other, placing them at the end of branches on a node of the tree, and replacing their distances from the other sequences by an average value.

The guide tree used by pileup and CLUSTAL should never be used to infer phylogeny! It has been derived from the distances between pairwise aligned sequences and these distances are not necessarily the same as the distances between sequence pairs taken from the multiple sequence alignment.

Go to: previous article - next article - Table to contents



NCBI

BLAST

BLAST Frequently Asked Questions (FAQ)

Tips and Hints:

Troubleshooting:

Tips and hints

Q: Which BLAST program should I use?

You have many choices to make between different BLAST programs and databases. Some of these choices are better for answering some questions then others. We have created a selection chart to help you make the decision of BLAST program for the question you are asking. This is the "BLAST Program Selection Guide".

Q: How can I search a batch of sequences with BLAST?

There are three options for "Batch" BLAST searches:
1) Web MegaBLAST EST analysis tool: This program is optimized for aligning nucleotide sequences that differ slightly as a result of sequencing or other similar "errors". MegaBLAST is good for scanning a large number of EST type sequences (about 500 kb in length) against large database in search of the closest matches. You can import a file EST sequences in FASTA format or as a list of GenBank accessions or/GIs and have them compared to the BLAST databases. The default is an easily reviewable Hit Table format, although you can download and save the results in Standard pairwise HTML or any of the other result output options. MegaBLAST is available from the BLAST web page, the standalone BLAST executables, or via the network BLAST client (see below).

2) Standalone BLAST executables:The Standalone BLAST executables are command line programs which run BLAST searches against local downloaded copies of the NCBI BLAST databases. The programs will handle either a single large file with multiple FASTA query sequences, or you can create a script to send multiple files one at a time. The executables are available for a wide variety of platforms, including many "flavors" of UNIX (LINUS, Solaris, etc.) Windows PC and even Mac OSX.

The Standalone executables are available at the anonymous FTP location: ftp://ftp.ncbi.nih.gov/blast/executables/ There is information on the Standalone BLAST executables available in the README file at ftp://ftp.ncbi.nih.gov/blast/documents/blast.txt which is also bundled with the downloaded binaries.

3) BLAST Network Client 'blastcl3':The BLAST 2.0 Network client will allow you to submit a single file of FASTA sequences over an internet connection to the NCBI BLAST databases. You submit searches through the client to the NCBI servers and do not need to download the database locally. The BLAST Network client executables are located at: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/ There are blastcl3 executables for various UNIX platforms, PC Windows and Macintosh.

Q: How can I write a program to submit jobs to NCBI's BLAST servers?

By using the URLAPI. Documentation also available in postscript and PDF.

Q: How can I limit my BLAST search based on Organism?

The option to limit a search to organism and even taxonomic classification is part of the "Limit by Entrez Search" option on most standard BLAST search pages. There is a pull down menu to select the most common organisms found in GenBank and also a field to input the species name, or classification (example: "eubacteria"). Using this option will cause your query sequences to be compared only to sequences in our databases from that organism.

There are also several "specialized" BLAST Pages devoted to different organisms on the main BLAST web page.

How can I limit my search to a subset of database sequences?

You can use the "Limit by Entrez Search" option found on most Standard BLASTR search pages to run an Entrez search and have your query sequence compared to the resutls of this search. For example, if you wanted to limit you search to all phosphorylase sequences from mouse you could enter the following valid Entrez search strategy in the Limit by Entrez field of the BLAST search page: phosphorylase AND "Mus musculus"[Organism]

Q: Is it possible to search for a motif or pattern with BLAST?

There are two general approaches to this type of questions. First do you wish to find if motifs exist in your query sequence, or do you have a known motif and wish to find other protiens or nucleotides with this motif?

In the first case, finding motifs in your query sequence can be done for proteins using the CDD (Conserved Domain Database) and CDART (Conserved Domain Architecture Retrieval Tool) tools. CDD allows you to compare your protein to an database of alignments and

profiles representing protein domains conserved in molecular evolution as well as 3-dimensional protein structures in the MMDB database. These tools use popular protien motif databases, PFam (http://pfam.wustl.edu/) and Smart (http://smart.embl-heidelberg.de) in addition to the MMDB database.

For conditions of the second case if you have a known motif and wish to identify other proteins with this motif you can use PHI-BLAST. PHI-BLAST searches take a motif pattern and protein sequence as input and then compares these to the NCBI protein databases looking for other proteins which contain conserved regions similar to the motif entered.

For nucleotides it is only possible to search with short query sequences representing your motif or region of interest with the Nucleotide BLAST "Search for short nearly exact matches" service from the main BLAST web page. This can find other sequences whicvh contain similar nucleotide patterns. however there are no database of nucleotide patterns which can identify patterns in your nucleotide query sequence.

You may also be interested in checking out other molecular biology web sites, such as those mentioned in the Other Molecular Biology Resources section at the end of this FAQ, for motif searching software.

Q: How do I perform a similarity search with a short peptide/nucleotide sequence?

There is a special page with pre-set parameters for searching with short sequences. You can access this page by clicking the "Search for short nearly exact matches" link on the main BLAST web page.

Essentially for these searches, the Expect value has been increased and the word size decreased to optimise for short hits which generally score a large E value require smaller word sizes to initiate formation of the HSP for extension. In addition, for proteins, the matix "PAM30" becomes the default which optimises hits to smaller sequences which have a lower percentage of evolutionary drift in general.

Q: Can I use BLAST to compare to two or more sequences in a multiple sequence alignment?

You can use the BLAST 2 Sequences service to compare two nucleotide or two protein sequences against each other using the Gapped BLAST algorithm. The this will allow you to perform a BLAST search between the two sequences allowing for the introduction of gaps (deletions and insertions) in the resulting alignment. Remember that BLAST is a "local" alignment program and does not make global alignments between sequences to calculate total percent homologies.

To compare one sequence against a specific sequence or set of sequences, you can also use a separate multiple sequence alignment program. There are many such software tools available to do this. You may also be interested in checking out other molecular biology web sites, such as those mentioned in the Other Molecular Biology Resources section at the end of this FAQ.

Q: What is the Expect (E) value?

The Expect value (E) is a parameter that describes the number of hits one can "expect" to

see just by chance when searching a database of a particular size. It decreases exponentially with the Score (S) that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. This means that the lower the E-value, or the closer it is to "0" the more "significant" the match is. However, keep in mind that searches with short sequences, can be virtually indentical and have relatively high EValue. This is because the calculation of the E-value also takes into account the length of the Query sequence. This is because shorter sequences have a high probability of occurring in the database purely by chance. For more details please see the calculations in the BLAST Course.

The Expect value can also be used as a convenient way to create a significance threshold for reporting results. You can change the Expect value threshold on most main BLAST search pages. When the Expect value is increased from the default value of 10, a larger list with more low-scoring hits can be reported.

Q: What is low-complexity sequence?

Regions with low-complexity sequence have an unusual composition andthis can create problems in sequence similarity searching (Wootton & Federhen, 1996). Low-complexity sequence can often be recognized by visual inspection. For example, the protein sequence PPCDPPPPPKDKKKKDDGPP has low complexity and so does the nucleotide sequence AAATAAAAAAAATAAAAAAT. Filters are used to remove low-complexity sequence because it can cause artifactual hits (please also see Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?)

In BLAST searches performed without a filter, often certain hits will be reported with high scores only because of the presence of a low-complexity region. Most often, this type of match cannot be thought of as the result of homology shared by the sequences. Rather, it is as if the low-complexity region is "sticky" and is pulling out many sequences that are not truly related.

Other Molecular Biology Resources:

The on-line BLAST Course was written by Dr. Stephen Altschul and discusses the basics of the Gapped BLAST algorithm. In addition the full text of the 1997 Nucleic Acids Research paper "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs" is also available on-line.

Other links:

European Bioinformatics Institute (EBI) BioCatalog
Indiana University IUBio Archive
Sequence manipulation site

Troubleshooting

Q: Causes for "No significant similarity found".

Below are several reasons that a BLAST search can result in the "No significant similarity found" message.

Short Sequences: There is a special BLAST optimized for searchig with small sequences.

AST FAQs

Go to tbe main BLAST web page and select the "Search for short nearly exact matches" link for Nucleotide – Nucleotide or Protein Protein sections.

Filtering: BLAST filters regions of low-complexity (for a description of low-complexity see "What is low-complexity sequence?" below). If your sequence contains large regions of "low complexity" it may not significant hits to the database. You can turn off filtering by setting the "Filter" option to "None" using the pull down tab.

Query Format: Another reason you may see the "No Significant Similarity found" message is using the wrong type of sequence in your search.

1) Accession/GI Number or FASTA. Check that you have the Input Data set to the correct format for your Query. Set the pull down menu to "Accession number or Gi" to search with GenBank accession numbers or Gi numbers. Set to FASTA for raw amino acid or nucleotide sequences. For more information on FASTA format, click here.

2) Sequence type and Program combination. You can search with an amino acid query sequence using the blastp and tblastn programs. With nucleotide query sequences you can use blastn, blastx, and tblastx. Please note that tblastx program cannot be used with the nr database on the BLAST Web page.

For more information on the BLAST programs, click here.

Q: Why does my search timeout on the BLAST servers?

Certain combinations of BLAST searches with large sequences against large databases can cause the BLAST servers to timeout. This has to do with a limit on the server CPU's which prevents sequences which generate many HSPs from hoarding server resources.

However there are some things you can do to prevent timeout and generate results from large sequences.

– Some sequences contain large regions of ALU repeats. In this case you can select the "Human Repeat" filtering option on the main BLAST search page. This will mask repeat regions which generate a large number of biologically uninteresting hits to the databases.

– Increase the Word Size to 20 – 25. With a default Word Size of 7, the BLAST algorithm finds initial HSPs of 7 bases in length and begins extension of these from either end. In a large sequence this can generate 100's of initial HSPs between the query sequence and even a single large genomic sequence in the databases. Increasing the Word Size to 25 makes the initial HSP smaller, limiting the number small initial fragments to be extended.

– Decrease the Expect value to 1.0 or lower. Many hits from large sequences are to many small fragments in the database. The expect value for these searches is such that decreasing the expect value will eliminate these results, and concentrate on results which are more likely to contain large coding regions and genomic fragments.

If you are still seeing a "timeout" error message after making the above changes, please contact blast-help@ncbi.nlm.nih.gov with the RID of your search.

Q: Why do I get the message "ERROR:BLASTSetUpSearch: Unable to calculate Karlin-Altschul params, check querysequence" ?

This will happen if your entire query sequence has been masked by low complexity filtering. You will need to turn filtering off to get hits. For further information on filtering, please read the sections of the BLAST FAQs on Q: What is low-complexity sequence? and also Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?

Q: Why do I get the message "ERROR: Blast: No valid letters to be indexed"?

You may have accidentally entered an accession number in the search box without changing the input selection from "Sequence in FASTA format" to "Accession or gi". You will also see this error message if too many ambiguity codes (R,Y,K,W,N, etc. fornucleotides) are present in your query sequence. Although BLAST allows ambiguity codes, be aware that these will always contribute a negative score in nucleic acid searches. Thus, sequences such as degenerate PCR primers with ambiguity codes maynot find any significant hits even though they may be designed from sequences that are present in the database.

Q: After running a search why do I see a string of "X"s (or "N"s) in my query sequence that I did not put there?

You are seeing the result of automatic filtering of your query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNN") or the letter "X" in protein sequences (e.g., "XXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment (Wootton amp; Federhen, 1996). Filter programs can eliminate these potentially confounding matches from the blast reports, leaving regions whose BLAST statistics reflect the specificity of their parities alignment. Queries searched with the blastn program are filtered with DUST. The other BLAST programs use SEG.

Q: How can I see low-similarity matches when there are many strong hits to my query sequence? Often, when the query is a member of a large sequence family, the summary hit list and the alignments returned only contain very high scoring hits. To look at low-similarity matches, you must increase the maximum number of results returned. On the BLAST Web pages, often it is sufficient to increase the size of the summary hit list and the number of alignments shown using the menus on the Advanced pages. However, it is possible to increase the lists even further using the Other Advanced Options box on the Advanced BLAST pages. For BLAST 2.0, "-v 2000", for example, will increase the number of descriptions returned in the summary hit list to 2000. The option "-b 2000" will similarly increase the number of alignments returned.

Q: I have heard that I will be penalized if I send a large number of sequences to the servers? .

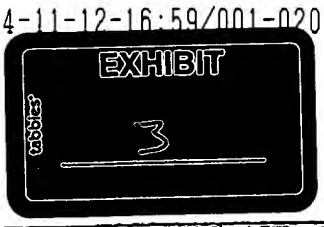
The NCBI WWW BLAST server is a shared resource and it would be unfair for a few users to monoplize it. To prevent this, the server keeps track of how many queries are in the queue for each user and penalzies those users with many queries in the queue. This is done by calculating a 'Time of Execution' (TOE). If a user has only one query in the queue, then the TOE is set to the current time. As a user adds more queries to the queue, then the TOE is set to the current time, plus 60 seconds for every query in the queue. An example would be if a user sent in five requests one after the other without waiting for any to be worked on, then the TOE's for the requests would be:

1st request: current time

LAST FAQs

2nd request: current time + 60 seconds
3rd request: current time + 120 seconds
4th request: current time + 180 seconds
5th request: current time + 240 seconds

The BLAST server works through requests in the order of earliest to latest TOE. A query will be executed before it's TOE, if there are no other queries with an earlier TOE. Users with large numbers of queries are encouraged to use the BLAST servers at off-peaks hours, which are from 8 p.m. to 8 a.m. (EST).



Reference 3

NCBI

More Information

glossary

ack to Education

LAST
iformation
itorial
uide

'SI-BLAST
utorial

Aore Information
imilarity searching
ules of thumb
lossary
eference list

Alignment
The process of lining up two or more sequences to achieve maximal levels of identity (and conservation, in the case of amino acid sequences) for the purpose of assessing the degree of similarity and the possibility of homology.

Algorithm
A fixed procedure embodied in a computer program.

Bioinformatics
The merger of biotechnology and information technology with the goal of revealing new insights and principles in biology.

Bit score
The value S' is derived from the raw alignment score S in which the statistical properties of the scoring system used have been taken into account. Because bit scores have been normalized with respect to the scoring system, they can be used to compare alignment scores from different searches.

BLAST
Basic Local Alignment Search Tool. (Altschul et al.) A sequence comparison algorithm optimized for speed used to search sequence databases for optimal local alignments to a query. The initial search is done for a word of length "W" that scores at least "T" when compared to the query using a substitution matrix. Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S". The "T" parameter dictates the speed and sensitivity of the search. For additional details, see one of the BLAST tutorials (Query or BLAST) or the narrative guide to BLAST.

BLOSUM
Blocks Substitution Matrix. A substitution matrix in which scores for each position are derived from *observations* of the frequencies of substitutions in blocks of local alignments in related proteins. Each matrix is tailored to a particular evolutionary distance. In the BLOSUM62 matrix, for example, the alignment from which scores were derived was created using sequences sharing no more than 62% identity. Sequences more identical than 62% are represented by a single sequence in the alignment so as to avoid over-weighting closely related family members. (Henikoff and Henikoff)

Conservation
Changes at a specific position of an amino acid or (less commonly, DNA) sequence that preserve the physico-chemical properties of the original residue.

Domain
A discrete portion of a protein assumed to fold independently of

the rest of the protein and possessing its own function.

DUST
A program for filtering low complexity regions from nucleic acid sequences.

E value
Expectation value. The number of different alignents with scores equivalent to or better than S that are expected to occur in a database search by chance. The lower the E value, the more significant the score.

FASTA
The first widely used algorithm for database similarity searching. The program looks for optimal local alignments by scanning the sequence for small matches called "words". Initially, the scores of segments in which there are multiple word hits are calculated ("init1"). Later the scores of several segments may be summed to generate an "initn" score. An optimized alignment that includes gaps is shown in the output as "opt". The sensitivity and speed of the search are inversely related and controlled by the "k-tup" variable which specifies the size of a "word". (Pearson and Lipman)

Filtering
Also known as Masking. The process of hiding regions of (nucleic acid or amino acid) sequence having characteristics that frequently lead to spurious high scores. See SEG and DUST.

gap
A space introduced into an alignment to compensate for insertions and deletions in one sequence relative to another. To prevent the accumulation of too many gaps in an alignment, introduction of a gap causes the deduction of a fixed amount (the gap score) from the alignment score. Extension of the gap to encompass additional nucleotides or amino acid is also penalized in the scoring of an alignment.

Global Alignment
The alignment of two nucleic acid or protein sequences over their entire length.

H
H is the relative entropy of the target and background residue frequencies. (Karlin and Altschul, 1990). H can be thought of as a measure of the average information (in bits) available per position that distinguishes an alignment from chance. At high values of H, short alignments can be distinguished by chance, whereas at lower H values, a longer alignment may be necessary. (Altschul, 1991)

Homology
Similarity attributed to descent from a common ancestor.

HSP
High-scoring segment pair. Local alignments with no gaps that achieve one of the top alignment scores in a given search.

Identity
The extent to which two (nucleotide or amino acid) sequences are invariant.

K
A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for search space size. The value K is used in converting a raw score (S) to a bit score (S').

lambda
A statistical parameter used in calculating BLAST scores that can be thought of as a natural scale for scoring system. The value lambda is used in converting a raw score (S) to a bit score (S').

Local Alignment
The alignment of some portion of two nucleic acid or protein sequences

Low Complexity Region (LCR)
Regions of biased composition including homopolymeric runs, short-period repeats, and more subtle overrepresentation of one or a few residues. The SEG program is used to mask or filter LCRs in amino acid queries. The DUST program is used to mask or filter LCRs in nucleic acid queries.

Masking
Also known as Filtering. The removal of repeated or low complexity regions from a sequence in order to improve the sensitivity of sequence similarity searches performed with that sequence.

Motif
A short conserved region in a protein sequence. Motifs are frequently highly conserved parts of domains.

Multiple Sequence Alignment
An alignment of three or more sequences with gaps inserted in the sequences such that residues with common structural positions and/or ancestral residues are aligned in the same column. Clustal W is one of the most widely used multiple sequence alignment programs

Optimal Alignment
An alignment of two sequences with the highest possible score.

Orthologous
Homologous sequences in different species that arose from a common ancestral gene during speciation; may or may not be responsible for a similar function.

P value
The probability of an alignment occurring with the score in question or better. The p value is calculated by relating the observed alignment score, S, to the expected distribution of HSP scores from comparisons of random sequences of the same length and composition as the query to the database. The most highly significant P values will

be those close to 0. P values and E values are different ways of representing the significance of the alignment.

PAM

Percent Accepted Mutation. A unit introduced by Dayhoff et al. to quantify the amount of evolutionary change in a protein sequence. 1.0 PAM unit, is the amount of evolution which will change, on average, 1% of amino acids in a protein sequence. A PAM(x) substitution matrix is a look-up table in which scores for each amino acid substitution have been calculated based on the frequency of that substitution in closely related proteins that have experienced a certain amount (x) of evolutionary divergence.

Paralogous

Homologous sequences within a single species that arose by gene duplication.

Profile

A table that lists the frequencies of each amino acid in each position of protein sequence. Frequencies are calculated from multiple alignments of sequences containing a domain of interest. See also PSSM.

Proteomics

Systematic analysis of protein expression of normal and diseased tissues that involves the separation, identification and characterization of all of the proteins in an organism.

PSI-BLAST

Position-Specific Iterative BLAST. An iterative search using the BLAST algorithm. A profile is built after the initial search, which is then used in subsequent searches. The process may be repeated, if desired with new sequences found in each cycle used to refine the profile. Details can be found in this discussion of PSI-BLAST. (Altschul et al.)

PSSM

Position-specific scoring matrix; see profile. The PSSM gives the log-odds score for finding a particular matching amino acid in a target sequence.

Query

The input sequence (or other type of search term) with which all of the entries in a database are to be compared.

Raw Score

The score of an alignment, S, calculated as the sum of substitution and gap scores. Substitution scores are given by a look-up table (see PAM, BLOSUM). Gap scores are typically calculated as the sum of G, the gap opening penalty and L, the gap extension penalty. For a gap of length n, the gap cost would be G+Ln. The choice of gap costs, G and L is empirical, but it is customary to choose a high value for G (10-15)and a low value for L (1-2).

Similarity

The extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. In BLAST similarity refers to a positive matrix score.

SEG

A program for filtering low complexity regions in amino acid sequences. Residues that have been masked are represented as "X" in an alignment. SEG filtering is performed by default in the blastp subroutine of BLAST 2.0. (Wootton and Federhen)

Substitution

The presence of a non-identical amino acid at a given position in an alignment. If the aligned residues have similar physico-chemical properties the substitution is said to be "conservative".

Substitution Matrix

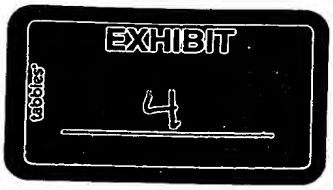
A substitution matrix containing values proportional to the probability that amino acid i mutates into amino acid j for all pairs of amino acids. such matrices are constructed by assembling a large and diverse sample of verified pairwise alignments of amino acids. If the sample is large enough to be statistically significant, the resulting matrices should reflect the true probabilities of mutations occurring through a period of evolution.

Unitary Matrix

Also known as Identity Matrix. A scoring system in which only identical characters receive a positive score.

Revised August 13, 2004

Eur. J. Biochem. 38, 103—110 (1973)



The Function of *myo*-Inositol in the Biosynthesis of Raffinose

Purification and Characterization of Galactinol:Sucrose 6-Galactosyltransferase from *Vicia faba* Seeds

Ludwig Lehle and Widmar Tanner

Fachbereich Biologie der Universität Regensburg

(Received April 18/June 28, 1973)

1. An enzyme from *Vicia faba* seeds is described which transfers the galactosyl moiety of galactinol to sucrose giving rise to raffinose and *myo*-inositol.

2. The enzyme was purified about 400-fold through 6 steps. A molecular weight of 80000 has been determined by gel-filtration and of 100000 by glycerol density gradient centrifugation.

3. The enzyme galactinol:sucrose 6-galactosyl transferase is different from α-galactosidase; these two activities as well as the stachyose-synthesizing enzyme separate during purification.

4. The transferase showed a high acceptor specificity. Out of 10 acceptors tested a transfer only to sucrose took place. This transfer was 5 times faster than the hydrolysis of galactinol. Galactinol, *p*-nitrophenyl-α-D-galactopyranoside and raffinose, but not UDP-galactose, could act as donors.

5. The enzyme catalyzes an exchange reaction between raffinose and [^{14}C]sucrose. This partial reaction is less sensitive towards heat inactivation and SH-poisons than the total reaction.

6. The pH-optimum of the reaction was found to be pH 7.0, the temperature optimum 42 °C. Heat inactivation could be prevented to some extent by galactinol and raffinose. In the presence of 0.4 mM sucrose the K_m-value for galactinol was 7 mM and for raffinose 10 mM. For sucrose a K_m-value of 1 mM in the synthesis reaction has been determined.

7. The transferase activity is high enough to explain the synthesis rate *in vivo* of all the raffinose-type sugars present in the seeds.

8. The physiological meaning of the results as well as the metabolic function of *myo*-inositol is discussed.

One of the major exceptions to Leloir's mechanism [1] of glycosidic linkage formation in nature has been discovered in the biosynthesis of a group of plant oligosaccharides, the sugars of the raffinose family [2,3]. Besides sucrose these sugars are the most common and widespread ones in higher plants and have a function as storage and transport material [4—6]. Whereas evidence *in vivo* and *in vitro* [2,7—9] has firmly established that the biosynthesis of stachyose and verbascose proceeds *via* a transglycosylation of the galactosyl-moiety from galactinol [L-1-(*O*-α-D-galactopyranosyl)-*myo*-inositol] to raffinose and stachyose, respectively [Eqns (3) and (4) below], conflicting evidence has been published concerning the biosynthesis of raffinose, the smallest member of the homologous series of these oligosaccharides.

On the one hand evidence for the reaction sequence (1) and (2), analogous to stachyose and verbascose synthesis, has been presented [10]. On the other hand a transfer of the galactosyl moiety from UDP-galactose to sucrose has

UDP-galactose + *myo*-inositol → galactinol + UDP (1)

Galactinol + sucrose ⇄ raffinose + *myo*-inositol (2)

Galactinol + raffinose ⇄ stachyose + *myo*-inositol (3)

Galactinol + stachyose ⇄ verbascose + *myo*-inositol (4)

Abbreviation. Gal-αONp, *p*-nitrophenyl-α-D-galactopyranoside.

Trivial Name. Galactinol, L-1-(*O*-α-D-galactopyranosyl)-*myo*-inositol.

Enzymes. α-Galactosidase or α-D-galactoside galactohydrolase (EC 3.2.1.22); galactinol : raffinose 6-galactosyltransferase (EC 2.4.1.–); aldolase or fructose-1,6-bisphosphate D-glyceraldehyde-3-phosphate lyase (EC 4.1.2.13).

been reported [11—13]. However, in this case the enzyme preparations were fairly crude and the possibility cannot be excluded that the sum of reaction (1) and (2) has been measured. Reaction (1) has been originally described by Frydman and Neufeld [14].

In the report to follow a 400-fold purification of the galactinol: sucrose 6-galactosyl transferase, the enzyme catalyzing reaction (2), from *Vicia faba* seeds will be described. The enzyme also catalyzes an exchange reaction between raffinose and sucrose, which is considerably more stable than the reaction responsible for net synthesis of raffinose. This latter observation explains the fact that Moreno an Cardini [15] have been able to observe only the exchange reaction in wheat germ extracts.

MATERIALS AND METHODS

Purification Procedure

All procedures were carried out at about 4 °C.

Step 1. Preparation of Crude Extract. 200 g ripe seeds from *Vicia faba* were powdered in a Waring Blendor and then extracted in a chilled mortar in two portions each with 200 ml of 0.1 M Tris-HCl buffer pH 7.3 containing dithioerythritol 5 mM. The homogenate was centrifuged for 30 min at 27000 × *g* giving a clear supernatant of about 250 ml.

Step 2. Treatment with Protamine Sulfate. The supernatant was brought to a protein concentration of 50 mg/ml with the same buffer as used for step 1. A 2% protamine sulfate solution was added to a final ratio of 9 mg protamine sulfate per 100 mg protein. After 30 min of stirring, the resulting precipitate was centrifuged off and discarded.

Step 3. Ammonium Sulfate Fractionation. To the protamine-treated supernatant saturated, cold ammonium sulfate solution, pH 7.3, was slowly added with continuous stirring to give 33% saturation. After 30 min, the precipitate was separated by centrifugation and the supernatant was brought to 55% saturation. The pellet obtained after centrifugation was dissolved in 70 ml 0.1 M Tris-HCl pH 7.3 containing 5 mM dithioerythritol and dialyzed overnight against 3 l of 0.05 M Tris-HCl pH 7.5, containing 1 mM dithioerythritol.

Step 4. Column Chromatography on DEAE-Cellulose. The dialyzed enzyme solution was adsorbed on a DEAE-cellulose column (2.5 × 30 cm) which had been equilibrated with 0.01 M Tris-HCl pH 7.5 containing 0.05 M KCl and 1 mM dithioerythritol. After the column was washed with equilibration buffer until all protein not bound was removed, 1 l linear gradient of 0.05 M KCl to 0.2 M KCl in 0.01 M Tris-HCl with 1 mM dithioerythritol was used for elution. Fractions of 6 ml were collected and those with the highest specific activity were pooled and concentrated to a small volume in an Aminco ultrafiltration cell with filter No XM-50.

Step 5. Sephadex G-200 Gel Chromatography. The pooled and concentrated fractions were loaded onto a column (2.5 × 80 cm) of Sephadex G-200, equilibrated with 0.01 M Tris-HCl buffer pH 7.5 containing 0.1 M KCl and 2 mM dithioerythritol. The column was eluted at a flow rate of 4 ml/h; 2-ml fractions were collected and the active fractions (100—120) were pooled and concentrated as described before.

Step 6. Hydroxyapatite Chromatography. After dialysis against 0.01 M Tris-HCl with 2 mM dithioerythritol pH 7.5, the enzyme solution was applied to a column (2.5 × 13 cm) of hydroxyapatite, which had been equilibrated with 0.01 M potassium phosphate buffer pH 7.5 containing 2 mM dithioerythritol. Elution was carried out stepwise with 100 ml potassium phosphate buffer of the following concentrations: (a) 0.01 M; (b) 0.05 M; (c) 0.1 M; (d) 0.2 M. The enzyme was eluted with 0.2 M buffer. The active fractions were again concentrated as described above.

Tests for Enzymic Activities

Galactosyltransferase: Synthesis and Exchange Reaction. Two tests have been used, to measure the transfer of the galactosyl moiety from galactinol to sucrose. In test I the amount of [^{14}C]raffinose formed from [^{14}C]sucrose has been determined. The incubation mixture contained in a total volume of 50 µl: 5 µmol Tris-HCl pH 7.2, 1 µmol galactinol, 0.02 µmol [^{14}C]sucrose (35 µCi/µmol) and enzyme. After incubation of 1—4 h at 32 °C the reaction was stopped with 0.2 ml ethanol and the preparation was centrifuged; the supernatant fluid was separated on Whatman No 1 in the solvent system *n*-butanol—pyridine—water—acetic acid (60:40:30:3, v/v/v/v). Radioactive spots were located with a strip scanner, cut out, and measured directly on paper in a scintillation counter in toluene—2,5-diphenyloxazole (efficiency 70%). This test was also applied for the exchange reaction with the only exception that 0.5 µmol raffinose was used instead of galactinol. The linear relationship between product formation, protein concentration up to 4 mg and incubation time up to 6 h has already been demonstrated for both reaction [10] and has since also been shown to be valid for the more purified enzyme preparations used in the kinetic experiments.

Test II is based on the galactosyl transfer from ^{14}C-labelled galactinol to sucrose. With this test one can study in addition the amount of galactose set free by the hydrolyzing activity of the transferase. The incubation mixture contained in a total volume of 50 µl: 5 µmol Tris-HCl pH 7.2, 0.013 µmol

[^{14}C]galactinol (7 μCi/μmol), 0.5 μmol sucrose and enzyme. The chromatographic separation was carried out in a solvent system of α-picoline—ammonia—water (70:28:2, v/v/v) until the front had reached half way down the paper. Then a second run in the solvent system *n*-butanol—pyridine—acetic acid—water (60:40:3:30, v/v/v/v) followed. Other conditions were the same as in test I.

α-Galactosidase. The enzyme was assayed by following the initial rate of *p*-nitrophenyl-α-D-galactopyranoside (Gal-αONp) hydrolysis. Enzyme solution was incubated at 32 °C with 25 μmol potassium phosphate buffer pH 5.5 and 6.0 μmol Gal-αONp for 15 min. The reaction was stopped by adding 5.0 ml of cold 0.1 M Na_2CO_3 and the yellow colour of *p*-nitrophenol was measured at 405 nm. Controls with Gal-αONp as well as with protein alone were run concurrently and all values appropriately corrected.

Determinations of Molecular Weight

The molecular weight was determined on a Sephadex G-200 column (2.5 × 80 cm) according to Andrews [16]. The column was eluted with 0.01 M Tris-HCl pH 7.5 containing 0.1 M KCl and 2 mM dithioerythritol. The calibration was obtained by determination of the elution volumes of a number of reference proteins of known molecular weight. The sedimentation constant of the enzyme was determined by centrifugation through a linear 5-ml gradient ranging from 5—20% glycerol in 0.05 M Tris-HCl pH 7.5 containing 5 mM dithioerythritol. The samples were centrifuged in the SWL 50 rotor of a Spinco L 2-65 B for 14 h at 0 °C. Then the tubes were punctured and fractions of 3 drops collected with the aid of a fraction collector. As reference protein aldolase was used.

Polyacrylamide-Gel Electrophoresis

The purity of the various purification steps was routinely checked by polyacrylamide gel electrophoresis in a 7.5% acrylamide gel according to Maurer [17]. Electrophoresis was performed at 2.0 mA/tube until the bromphenolblue band had reached the bottom of the tube. Fixation and staining were carried out according to Chrambach *et al.* [18].

Other Procedures

Protein determinations were carried out according to Lowry *et al.* [19] with bovine serum albumin as a standard. Labelled galactinol was isolated by paper chromatography from the water-soluble extract of lamium leaves after photosynthesis in $^{14}CO_2$ according to Kandler [20]. A sample of unlabelled galactinol was generously supplied by Dr R. M. McCready (USDA, Agricultural Research Service, Albany).

RESULTS

Purification of Galactinol: Sucrose 6-Galactosyltransferase

Table 1 summarizes the results of the overall purification. Starting from a crude extract which a specific activity of 0.071 nmol × mg^{-1} × h^{-1} a preparation was obtained with a specific activity of 29.8 nmol × mg^{-1} × h^{-1} (peak II of hydroxyapatite chromatography). The results show that the enzyme catalyzing the synthesis of stachyose [8] separates

Table 1. *Purification procedure for galactinol : sucrose 6-galactosyltransferase*
Figures in brackets represent percentage of original activity. Stachyose was measured as described previously [7]

Fraction	Total protein	Raffinose synthesis		Exchange reaction		Stachyose synthesis		α-Galactosidase activity	
		Total activity	Specific activity	Total activity	Specific activity	Total activity	Specific activity	Total activity	Specific activity
	mg	nmol/h	nmol × h^{-1} × mg^{-1}	nmol/h	nmol × h^{-1} × mg^{-1}	nmol/h	nmol × h^{-1} × mg^{-1}	μmol/h	μmol × h^{-1} × mg^{-1}
1. Crude extract	22000	1576 (100)	0.071	1980 (100)	0.091	2148 (100)	0.098	1540 (100)	70
2. Protamine sulfate	9350	1750 (110)	0.187	2072 (105)	0.222	2306 (107)	0.246	272 (18)	29
3. Ammonium sulfate	2688	1398 (88)	0.520	1730 (88)	0.647	1488 (69)	0.553	161 (10)	58
4. DEAE-cellulose	182	557 (35)	3.007	614 (31)	3.376	0 (0)	0	5 (0.3)	29
5. Sephadex G-200	50	262 (17)	5.273	359 (18)	7.168	0 (0)	0	0 (0)	0
6. Hydroxy-apatite									
Peak I	4.5	3.1 (0.2)	0.703	32.9 (1.6)	7.326	0 (0)	0	0 (0)	0
Peak II	2.3	68.5 (4.3)	29.800	58.5 (2.9)	25.345	0 (0)	0	0 (0)	0

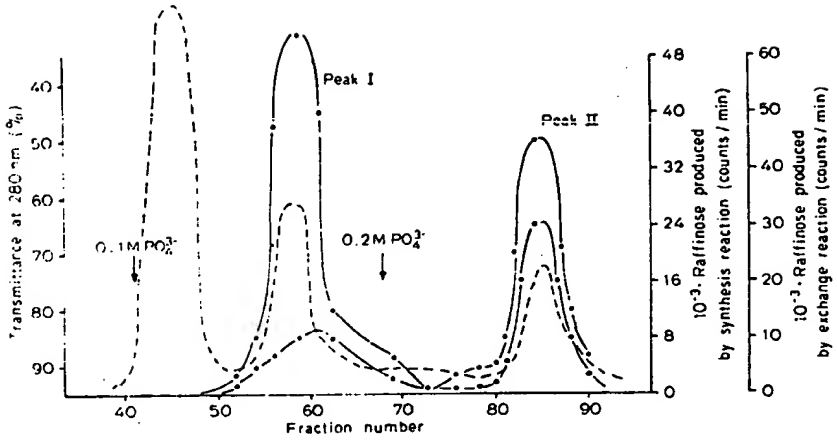


Fig. 1. *Hydroxyapatite chromatography of galactinol : sucrose 6-galactosyltransferase.* Suitable aliquots were tested by test I for synthesis activity (●——●) and exchange activity (○- - - -○); absorbance at 280 nm (----)

from the corresponding raffinose-synthesizing enzyme In addition it has to be pointed out that the purified enzyme is different from an α-galactosidase, since the hydrolyzing activity towards *p*-nitrophenyl-α-D-galactopyranoside (Gal-αONp), known to be a good substrate for α-galactosidases, separates likewise from the raffinose-synthesizing activity. Since the galactinol : sucrose 6-galactosyl transferase is the most labile of the plant galactosyl transferases known (*e.g.* [10]), it seems unlikely that an inactivation instead of a separation of the other two enzymes had occurred during the purification. The considerable decrease of the α-galactosidase activity in step 2 may on the other hand be the reason for the observed increase of the total raffinose-synthesizing activity in this fraction, since less of the newly synthesized raffinose will be lost by hydrolysis.

The preparation from *Vicia faba* also catalyzes an exchange reaction between raffinose and sucrose according to the following equation:

Raffinose + [^{14}C]sucrose $\rightleftarrows$ [^{14}C]raffinose + sucrose.

This reaction has originally been described by Moreno and Cardini [15]: their enzyme preparation from wheat germ, however, did not catalyze the synthesis of raffinose. Through all the steps given in Table 1 (except for step 6; see Discussion below) the exchange reaction parallels the synthesis activity. Thus both reactions most likely are catalyzed by one and the same enzme.

In the last purification step two active transferase peaks (I and II in Fig. 1) were obtained. The main fraction, peak II, was eluted with a buffer concentration of 0.2 M. Peak I, which had much lower specific activity, appeared at 0.1 M. Both fractions were able to catalyze the synthesis as well as the exchange reaction, although at different relative rates. Whereas peak I catalyzes the exchange reaction about 10 times faster than the synthesis of raffinose, peak II catalyzes the exchange reaction only at 85% the rate of synthesis reaction. Further experiments indicated that peak I is a modified form of the enzyme, which has lost most of its raffinose-synthesizing activity and shows a different elution behaviour as compared to the native enzyme. Thus, when peak II was chromatographed a second time on hydroxyapatite, again an active peak I and II was obtained. The observation made previously, that the activity for raffinose synthesis is lost more readily than the activity of the exchange reaction [10], is in agreement with the above finding.

When checked for purity by polyacrylamide gel electrophoresis the 400-fold purified fraction was not yet homogeneous; one major and three minor bands have been observed (Fig. 2). Although a strong attempt has been made to correlate the enzyme activity with one of the bands, this has failed; the enzymic activity always got lost during gel-electrophoreses, even in the presence of a variety of protecting agents.

The enzyme remained in the supernatant when the enzyme solution was centrifuged at 100000 × *g* for 1 h.

Determination of Molecular Weight

The molecular weight of the enzyme was determined by two different methods. From the sedimentation profile in a glycerol density gradient a molecular weight of 100000 was obtained when compared to the sedimentation of aldolase (Fig. 3). With Sephadex G-200 gel chromatography on a standardized column (Fig. 4) a value of 80000 was

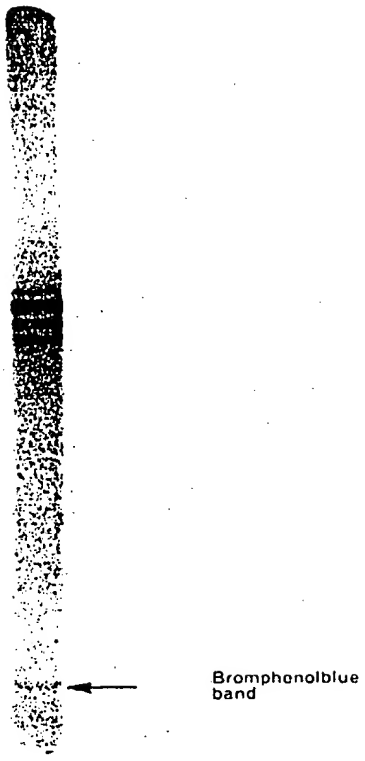


Fig. 2. *Disc-gel electrophoresis of peak II of the hydroxyapatite column*

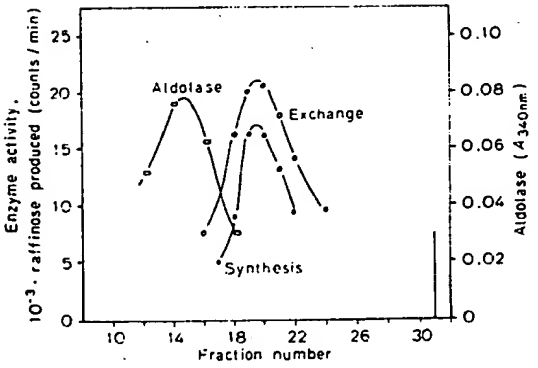


Fig. 3. *Sedimentation profile of galactinol : sucrose 6-galactosyltransferase in a 5—20% glycerol density gradient.* 100 µg of purified enzyme (Sephadex fraction) and 500 µg of aldolase were centrifuged for 16 h at 40000 rev./min. Enzyme activity has been tested for synthesis reaction (●——●) and exchange reaction (O——O). Aldolase served as a marker

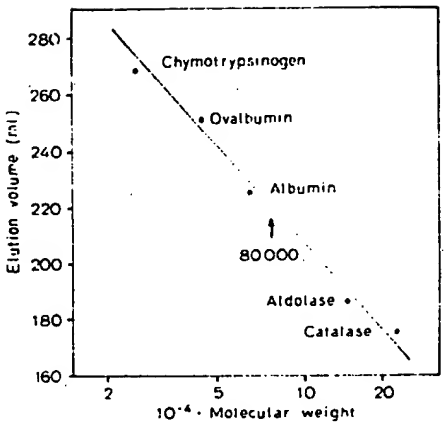


Fig. 4. *Determination of the molecular weight by Sephadex G-200 gel filtration.* The procedure is described in the text. The arrow indicates the position of the enzyme. Molecular weight is plotted on a log scale

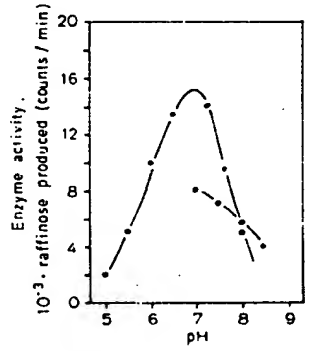


Fig. 5. *pH dependence of galactinol : sucrose 6-galactosyltransferase.* Assays were performed with potassium phosphate buffer (O - - - O) and Tris-HCl buffer (● - - - ●)

determined. In each case, however, the same values were observed, whether synthesis or exchange activity had been tested.

Stability

When stored at 4 °C the crude extract lost 50% of its original activity in the synthesis reaction and 30% in exchange reaction within 3 days. The activity of the purified enzyme when frozen was unchanged for at least a month.

pH Optimum

The enzyme showed an optimum around pH 7.0. In the presence of potassium phosphate buffer the activity was higher than in the presence of Tris-HCl buffer (Fig. 5).

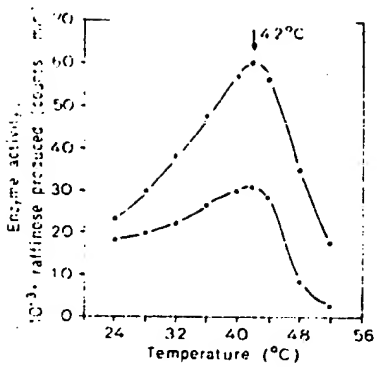


Fig. 6. *Effect of temperature on the activity of galactinol : sucrose 6-galactosyltransferase (Sephadex fraction).* (● ··· ●) Synthesis reaction; (○ ··· ○) exchange reaction

Table 2. *Effect of substrates on heat inactivation of galactinol : sucrose 6-galactosyltransferase*
140 µg enzyme (Sephadex fraction) was preincubated with or without substrate at 50 °C for 10 min. Then the rest of the incubation mixture was added and the test was carried out under standard conditions for 150 min

Substrate	Enzyme activity after preincubation			
	Synthesis reaction		Exchange reaction	
	counts/min	%	counts/min	%
Control (without heat)	22755	100	46258	100
− substrates	4350	19	16610	36
+ donor[a]	7507	33	29605	64
+ acceptor[b]	4515	20	17994	38

[a] Galactinol and raffinose.
[b] Sucrose.

Effect of Temperature on the Enzyme Activity

Fig. 6 shows the temperature profile of the enzyme activities. Maximum rate for both reactions occurs at 42 °C with a sharp drop beyond 44 °C, the synthesis reactions being somewhat more sensitive than the exchange reaction. In this connection it has been observed that galactional and raffinose prevent to some extent inactivation by heat (Table 2). Sucrose, however, at the concentrations used had no effect.

Inhibition with Sulfhydryl-Specific Reagents

One of the main reasons that the first step in the biosynthesis of raffinose sugars escaped detection for a rather long time has certainly been the requirement of the enzyme for strong SH-protecting agents [10]. This is especially true for the synthesis reaction. The different susceptibility of synthesis and exchange reaction is also reflected by the inhibition of the enzyme with iodoacetamide and *N*-ethylmaleimide (Table 3). The heavy metal ions Ag^+, Hg^{2+}, Zn^{2+} and Al^{3+} at a concentration of 1 mM inhibited the synthesis reaction of the enzyme to 100%; Mn^{2+} inhibited to 60%.

Table 3. *Inhibition of synthesis and exchange reaction by thiol-group specific reagents*
150 µg enzyme (Sephadex fraction) was incubated under standard conditions for 2 h. The inhibitor concentration was 1 mM

Inhibitor	Synthesis reaction		Exchange reaction	
	counts/min	(%)	counts/min	(%)
Control	21510	(0)	35500	(0)
Iodoacetamide	10995	(49)	31510	(11)
N-Ethyl-maleimide	2153	(90)	25490	(16)

Enzyme Kinetics

K_m values for galactinol, sucrose and raffinose have been determined (Fig. 7, Fig. 8 and Table 4). The Michaelis constant for sucrose was found to be 1 mM in the synthesis reaction and 2.9 mM in the exchange reaction in the presence of 0.02 M galactinol and raffinose, respectively. When the galactinol and raffinose concentrations were decreased 100-fold, the K_m for sucrose in the synthesis reaction stayed the same (1.4 mM). It was, however, considerably lower (0.47 mM) in the exchange experiment. This is consistent with the assumption that the binding site for raffinose and sucrose might be identical; a high raffinose concentration would then act as competitive inhibitor. On the other hand the sites for galactinol and sucrose seem to be different; a change in the concentration of galactinol has no influence on the K_m of sucrose. It has to be pointed out that the K_m-values for galactinol and raffinose given in Table 4 are only valid for a sucrose concentration of 0.4 mM.

Acceptor and Donor Specificity

The acceptor specificity has been tested by measuring the transfer of the ^{14}C-labelled galactosyl moiety from [^{14}C]galactinol to various acceptors. Out of 10 acceptors tested only a transfer to sucrose could be observed (Table 5). The purified enzyme cannot catalyze the biosynthesis of stachyose and verbascose. Both these enzymic activities have already been found in seeds from *Vicia faba* [8]. It should be noted that during the incubation of [^{14}C]galactinol some free [^{14}C]galactose was obtained due to the hydrolysis of galactinol. However, in the presence of sucrose the amount of galactose transferred was nearly 5 times greater than the amount of galactinol hydrolyzed (Table 5). In the absence of any acceptor considerable more galactose was set free. This can be interpreted as a competitition of sucrose with water. As donors only galactinol,

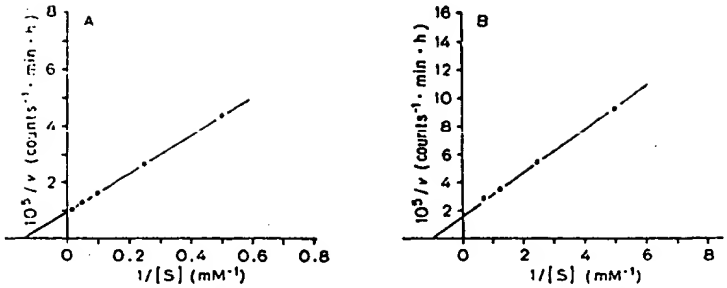


Fig. 7. *Lineweaver-Burk plots: raffinose-synthesis reaction.* (A) Galactinol in the presence of 0.4 mM sucrose; (B) sucrose in the presence of 0.01 M galactinol

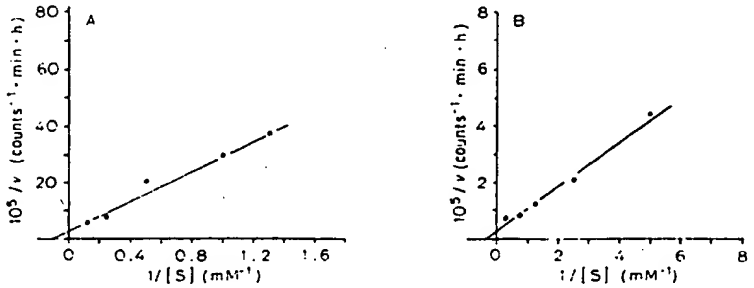


Fig. 8. *Lineweaver-Burk plots: exchange reaction.* (A) Raffinose in the presence of 0.4 mM sucrose; (B) sucrose in the presence of 0.01 M raffinose

Table 4. *K_m-values of galactinol : sucrose 6-galactosyltransferase*

Substrate	K_m value of	
	Synthesis reaction	Exchange reaction
	mM	mM
Galactinol	7.0	—
Raffinose	—	10
Sucrose[a]	1.0	2.9
Sucrose[b]	1.4	0.47

[a] In the presence of 0.02 M donor (*i.e.* galactinol and raffinose, respectively).
[b] In the presence of 0.2 mM donor.

Table 5. *Acceptor and donor specificity of galactinol:sucrose 6-galactosyltransferase (Sephadex fraction)*
In the acceptor experiment the incubation mixture contained in a total volume of 50 µl: 5 µmol Tris-HCl pH 7.2, 0.5 µmol acceptor, 0.039 µmol [^{14}C]galactinol (7 µCi/µmol) and 0.3 mg protein. After 4 h at 32 °C the reaction was stopped. In the donor experiment the incubation mixture contained 0.5 µmol donor, 0.02 µmol [^{14}C]sucrose (35 µCi/µmol), 5 µmol Tris-HCl and 0.1 mg protein. The incubation time was 1 h at 32 °C. Raffinose, stachyose, fructose, glucose, galactose, lactose, cellobiose, melibiose and glycerol do not act as acceptors

Donor	Acceptor	Reaction product	
		Raffinose	Galactose
		counts/min	
[^{14}C]Galactinol	Sucrose	11505	2495
[^{14}C]Galactinol	H_2O	0	10014
Galactinol	[^{14}C]Sucrose	24012	
UDP-Gal		0	
Gal-αONp		13210	
Raffinose		33405	
Stachyose		1005	
Melibiose		995	

Gal-αONp (an unphysiological substrate) and raffinose, *i.e.* in the exchange reaction, work to a significant extent (Table 5). Transfer from UDP-galactose to sucrose has been observed neither with the purified enzyme nor the crude extract [10].

DISCUSSION

The enzyme catalyzing the transfer of the galactosyl moiety from galactinol to sucrose has been isolated, purified and characterized. The results indicate that the enzyme is clearly different from any of the α-galactosidases described [5, 21—25]. Thus the hydrolyzing activity towards Gal-αONp, a typical substrate for α-galactosidases, separates from the raffinose-synthesizing enzyme during the purification. Furthermore the high substrate specificity as well as the efficiency of the transfer have to be pointed out, when the enzyme is compared with

α-galactosidases. It is proposed to call the enzyme galactinol : sucrose 6-galactosyl-transferase and to group it among the glycosyl transferases.

The exchange reaction is catalyzed by the same enzyme which is responsible for raffinose synthesis. This has also been expected in analogy to similar transfer reactions [7,26,27].

The enzyme activity of 7.9 nmol raffinose formed $\times h^{-1} \times g$ seeds^{-1} (Table 1) corresponds to an activity of 28.4 nmol $\times h^{-1} \times g$ seeds^{-1} at the physiological sucrose concentration of 10 mM. This rate is high enough to explain the synthesis rate *in vivo* for raffinose and for all the other higher homologues of the raffinose sugars during the ripening period. Thus the enzyme is able to synthesize 2.5 μmol raffinose, the amount actually present in 1 g of seeds, in less than 4 days.

The synthesis of the total amount of the other raffinose-type sugars (21.4 μmol/g seed) would take about one month, which corresponds reasonably well to the ripening period of the seeds.

In addition the results of the biosynthesis of raffinose and its higher homologues *in vitro* with respect to the function of galactinol are in agreement with the studies *in vivo* by Senser and Kandler [2,29]. It seems without doubt now, that the biosynthesis of all the raffinose sugars proceeds *via* galactinol. The physiological meaning of the detour taken by the galactosyl moiety is not understood at present. Perhaps it has to be seen in relation to the observation that *myo*-inositol and galactinol inhibit α-galactosidases, enzymes responsible for the decomposition of raffinose sugars [9,30].

Myo-inositol has been known as a growth factor for yeasts and many tissue cultures [31—33] for a long time. Since these cells do not contain sugars of the raffinose family the cofactor-like role, which *myo*-inositol plays in the biosynthesis of oligosaccharides, cannot explain its function as a growth factor. It seems likely, however, that *myo*-inositol is absolutely required in the form of phosphatidylinositols, which seem to be indispensable membrane components [34]. This is supported by the finding that transport mechanisms are impaired when cells lack *myo*-inositol [35—37].

We would like to thank Drs A. Böck and H. Kosakowski for helpful suggestions and advice.

REFERENCES

1. Cardini, C. E., Leloir, L. F. & Chiriboga, J. (1955) *J. Biol. Chem. 214*, 149—155.
2. Senser, M. & Kandler, O. (1967) *Z. Pflanzenphysiol. 57*, 376—388.
3. Tanner, W. & Kandler, O. (1966) *Plant. Physiol. 41*, 1540—1542.
4. French, D. (1954) *Adv. Carbohydr. Chem. 9*, 149—184.
5. Courtois, J. E. (1959) *Carbohydrate Chemistry of Substances of Biological Interest* (Wolfram, M. L., ed.) pp. 140—159, Pergamon Press, London.
6. Jeremias, K. (1963) *Bot. Stud. 15*, 1—96.
7. Tanner, W. & Kandler, O. (1968) *Eur. J. Biochem. 4*, 233—239.
8. Tanner, W., Lehle, L. & Kandler, O. (1967) *Biochem. Biophys. Res. Commun. 29*, 166—171.
9. Tanner, W. (1969) *Ann. N.Y. Acad. Sci. 165*, 726—742.
10. Lehle, L., Tanner, W. & Kandler, O. (1970) *Hoppe-Seyler's Z. Physiol. Chem. 351*, 1494—1498.
11. Bourne, E. J., Walter, M. W. & Pridham, J. B. (1965) *Biochem. J. 97*, 802—806.
12. Pridham, J. B. & Hassid, W. Z. (1965) *Plant. Physiol. 40*, 984—986.
13. Gomyo, T. & Nakamura, M. (1966) *Agric. Biol. Chem. 30*, 425—427.
14. Frydman, R. B. & Neufeld, E. F. (1963) *Biochem. Biophys. Res. Commun. 12*, 121—125.
15. Moreno, A. & Cardini, C. E. (1966) *Plant Physiol. 41*, 909—910.
16. Andrews, P. (1964) *Biochem. J. 91*, 222—233.
17. Maurer, H. R. (1968) *Disk Electrophorese*, pp. 39—47, Walter de Gruyter and Co, Berlin.
18. Chrambach, A., Reisfeld, R. A., Wyckhoff, M. & Zaccari, J. (1967) *Anal. Biochem. 20*, 150—154.
19. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265—275.
20. Kandler, O. (1964) *Ber. Dtsch. Bot. Ges. 77*, 62—73.
21. Courtois, J. E. & Petek, F. (1966) *Methods Enzymol. 8*, 565—571.
22. Courtois, J. E. & Petek, F. (1957) *Bull. Soc. Chim. Biol. 39*, 715.
23. Dey, P. M. & Pridham, J. B. (1969) *Biochem. J. 115*, 47—54.
24. Malhotra, O. P. & Dey, P. M. (1967) *Biochem. J. 103*, 739—743.
25. Yu-Te Li & Shetlar, M. R. (1964) *Arch. Biochem. Biophys. 108*, 301—313.
26. Glaser, L. (1964) *Compr. Biochem. 15*, 93—137.
27. Nisizawa, K. & Hashimoto, J. (1970) *The Carbohydrates* (Pigman, W. & Horton, D., eds) vol. IIA, pp. 241—300, Academic Press, New York.
28. Tanner, W., Seifarth, H. & Kandler, O. (1968) *Z. Pflanzenphysiol. 58*, 369—377.
29. Senser, M. & Kandler, O. (1967) *Phytochemistry, 6*, 1533—1540.
30. Sharma, Ch. B. (1971) *Biochem. Biophys. Res. Commun. 43*, 572—579.
31. Burkholder, P. R., McVeigh, I. & Meyer, D. (1944) *J. Bacteriol. 48*, 385—391.
32. Eagle, H., Oyama, V. I., Levy, M. & Freeman, A. E. (1957) *J. Biol. Chem. 226*, 191—205.
33. Wolter, K. E. & Skoog, F. (1966) *Am. J. Bot. 53*, 263—269.
34. Jung, P., Tanner, W. & Wolter, K. (1972) *Phytochemistry, 11*, 1655—1659.
35. Lembach, K. & Charalampous, F. C. (1967) *J. Biol. Chem. 242*, 2599—2605.
36. Charalampous, F. C. (1969) *J. Biol. Chem. 244*, 1705—1710.
37. Scarborough, G. A. (1971) *Biochem. Biophys. Res. Commun. 43*, 968—975.

L. Lehle and W. Tanner, Fachbereich Biologie der Universität, D-8400 Regensburg, Universitätsstraße 31, Federal Republic of Germany